Track 4: Case Study

Background

SignX Inc., a biopharmaceutical company is developing a novel therapeutic (Drug X) for the treatment of a solid tumor. The tumor tissue has been shown to have a characteristic alteration in gene expression when compared to normal tissue. This "signature" is detectable as early as two weeks after the onset of oncogenesis and remains constant over the first few weeks of tumor growth. In advanced stages, different signatures are found that are likely compounded by other pathophysiologic events associated with late stage disease.

The company has decided to explore if (1) a characteristic signature can be found in peripheral blood mononuclear cells (PBMCs) as a surrogate tissue, since they provide an advantage of having easier accessibility by less invasive means, and (2) the signature can be used as a measurement of therapeutic effect (efficacy biomarker). In order to discuss this exploratory approach of identifying and validating a gene signature biomarker, SignX Inc. decided to submit the data to the FDA as a voluntary genomic data submission (VGDS).

1. Submission of Pharmacogenomic Data to the FDA

From a phase 2 study that included 200 cancer patients and 200 controls, RNA was isolated from PBMCs and run on Affymetrix oligonucleotide microarrays. The cancer patients were followed over a period of 24 weeks and PBMCs were collected at 8 week intervals (0, 8, 16, and 24 weeks). 50 patients were treated with the "best of care" and did not receive therapy (Placebo group). The remaining 150 cancer patients were treated with Drug X. Together, a total of 1000 hybridizations were performed (4x (150 treated + 50 Placebo) cancer patients, 1x 200 controls). SignX Inc. performed a preliminary analysis of the data and decided to submit the data and analysis results to the FDA as a VGDS.

Questions:

- What data should be submitted?
    - raw data (*cel file, probe set file, image data)*
    - normalization algorithm
    - list of genes
    - biological interpretation of the data
    - MIAME guidelines
    - phenotypic information
- How can we link the submitted microarray data back to phenotypic data: what types of databases are needed and do they exist (i.e. gene expression signature to phenotype relationship)?
- What other guidelines currently under development should be followed to capture both microarray and phenotypic data?
    - HL-7/CDISC
    - MAQC
    - ERCC

2. Analysis of Pharmacogenomic Data

After the IPRG received the raw data, the data set was loaded into the ArrayTrack database and analysis system. A statistical analysis was carried out to identify differentially regulated genes in the PBMCs of cancer patients at baseline (200) prior to treatment versus normal, healthy individuals (200). In addition, gene expression signatures of treated versus untreated patients were obtained at the four different time points. This type of analysis was also performed at the FDA to reconstruct the analysis performed by the sponsor as well as to extend the knowledge generated by the interpretation of the data. However, the sponsor used a different analysis platform and different statistical approach to derive the gene signatures. Interestingly, both, FDA and sponsor were able to identify statistically relevant signatures, but the overlap of the gene lists between FDA and sponsor was varying in different cases.

Questions:

- What are the best practices for the normalization of hybridization data?
- What are the best practices for the statistical analysis of normalized hybridization data for changes in gene expression relative to control samples?
    - Stringency
    - P-value vs. fold change
    - False Discovery Rate and other statistical tools
- What are the best practices for the statistical analysis of normalized hybridization data for signatures associated with baseline expression levels?
    - Supervised learning methods
        - Consensus methods
        - Single classification models
    - Unsupervised learning methods

3. Biological Interpretation of Pharmacogenomic Data

After the data analysis was performed, the gene signatures were further evaluated for biological interpretation. The analysis was performed using KEGG, PathArt and the Ingenuity Pathways Analyses platforms and related information researched in PubMed and SafeBase**.** These biological pathway sets are used to:

1) compare the reconstruction analysis at the FDA with the analysis submitted by the sponsor.
2) generate additional useful information from the data consistent with interpretations generated by the sponsor and the FDA.

The comparison between patients and healthy volunteer samples did not find any interesting pathways; however, the analysis of treated versus untreated patients showed an activation of apoptosis and of chemokine signaling pathways in patients treated for more than 8 weeks.

Questions:

- Are the tools used sufficient to get biological interpretation?
- What other tools (such as GeneGo, BioCarta, KeyMolnet), should be used?
- How should the biological interpretation of gene expression data be used in a regulatory context?
- Based on this data, is it reasonable to expect that gene expression analysis of PBMCs represents gene expression changes in oncology?
    - Sigantures to predict time to death (TTD) and time to progression (TTP)
    - Signatures to predict drug response.
    - What other therapeutic areas do we know about where PBMCs are good surrogate reporter cells?
- Should data analysis tools and databases (knowledge bases) be integrated?

4. A Future Scenario: What should the ultimate "pharmacogenomics suite" look like?

It should cover early data capturing (clinical phenotypic data, but also analytical and experimental data), link to data analysis tools and databases and use these datasets for immediate visualization of relevant pathways and other useful biological information. It also should link to other types of "-omics" data such as proteomic and metabolomic data and to large compound specific databases to assess ASRs, into adverse event databases, and other resources such as the HapMap and SNP databases in order to extract the most relevant and up-to-date information/interpretation.

- Can we capture all this in a single application or is there a need for independent, but linked platforms?
- Can the analysis and interpretation be automated?
- How can the System approach play a regulatory role?